

Soft Computing Techniques in Spatial Databases

Markus Schneider

University of Florida

Department of Computer & Information Science & Engineering

Gainesville, FL 32611

USA

ABSTRACT

Spatial database systems and geographical information systems are currently only able to support geographical applications that deal with *crisp spatial objects*, that is, objects whose extent, shape, and boundary are precisely determined. Examples are land parcels, school districts, and state territories. However, many new, emerging applications are interested in modeling and processing geographic data that are inherently characterized by *spatial vagueness* or *spatial indeterminacy*. Examples are air polluted areas, temperature zones, and lakes. These applications require novel concepts due to the lack of adequate approaches and systems. In this chapter, we show how soft computing techniques can provide a solution to this problem. We give an overview of two type systems or algebras that can be integrated into database systems and utilized for the modeling and handling of spatial vagueness. The first type system, called *Vague Spatial Algebra (VASA)*, is based on well known, general, and exact models of *crisp spatial data types* and introduces *vague points*, *vague lines*, and *vague regions*. This enables an exact definition of the vague spatial data model since we can build it upon an already existing theory of spatial data types. The second type system, called *Fuzzy Spatial Algebra (FUSA)*, leverages fuzzy set theory and fuzzy topology and introduces novel *fuzzy spatial data types* for *fuzzy points*, *fuzzy lines*, and *fuzzy regions*. This enables an even more fine-grained modeling of spatial objects that do not have sharp boundaries and interiors or whose boundaries and interiors cannot be precisely determined. This chapter provides a formal definition of the structure and semantics of both type systems. Further, we introduce spatial set operations for both algebras and obtain vague and fuzzy versions of geometric intersection, union, and difference. Finally, we describe how these data types can be embedded into extensible databases and show some example queries.

INTRODUCTION

Spatial database systems (SDBS) are full-fledged database systems which, in addition to the functionality of standard database systems for alphanumeric data, provide special support for the storage, retrieval, management, and querying of spatial data, that is, objects in space. In particular, SDBS are used as the data management foundation of *Geographic Information Systems (GIS)*. In the literature, the common consensus prevails that special data types are necessary to adequately model geometry and to efficiently represent geometric data in database systems. These data types are commonly denoted as *spatial data types* (Schneider 1997) such as *point*, *line*, and *region*. We speak of *spatial objects* as instances of these data types. So far, the mapping of spatial phenomena of the real world leads almost exclusively to precisely defined spatial objects. Spatial data modeling implicitly assumes that the positions of points, the locations and routes of lines, and the extent and hence the boundary of regions are precisely determined

and universally recognized. This leads to *exact object models*. Examples are especially man-made spatial objects representing engineered artifacts (like monuments, highways, buildings, bridges) and predominantly immaterial spatial objects exerting social control (like countries, districts, and land parcels with their political, administrative, and cadastral boundaries). We denote this kind of entities as *crisp* or *determinate spatial objects*.

But for many geometric applications, the mapping into crisp spatial objects is an insufficient abstraction process since many geographic objects show the inherent feature of *spatial vagueness* or *spatial indeterminacy* (Burrough & Frank 1996). Current GIS and spatial database systems are not capable of supporting applications based on vague geometric data. In these applications, the positions of points are not exactly known, the locations and routes of lines are unclear, and regions do not have sharp boundaries, or their boundaries cannot be precisely determined. Examples are natural phenomena (like soil quality, vegetation, oceans, valleys, mountains, oil fields, biotopes, deserts, clouds, temperature zones, air pressure, sandbanks), cultural phenomena (like a Rhaeto-Romanic language speaking area) and social phenomena (like population density, unemployment rate, air pollution emission, terrorists' refuges and escape routes). We denote this kind of entities as *vague* or *indeterminate spatial objects*.

This chapter shows that and how different soft computing techniques can be leveraged to represent spatial vagueness. It gives an overview of two different type systems that can be integrated into database systems and whose types can be employed as attribute types in the same way as standard data types like *integer* or *string*. A fundamental design concept is that these new types are *not* represented through a database model (for example, relational, object-oriented, complex data model) but through *abstract data types* that encapsulate and hide the internal complexity of their values. This implies that a database model and its underlying theoretical framework (for example, relational database theory) does not have to be modified in any manner. The first type system, called *Vague Spatial Algebra (VASA)*, is based on well known, general, and exact models of *crisp spatial data types* and introduces *vague points*, *vague lines*, and *vague regions*. This enables an exact definition of the vague spatial data model since we can build it upon an already existing theory of spatial data types. The second type system, called *Fuzzy Spatial Algebra (FUSA)*, leverages fuzzy set theory and fuzzy topology and introduces novel *fuzzy spatial data types* for *fuzzy points*, *fuzzy lines*, and *fuzzy regions*. This enables an even more fine-grained modeling of spatial objects that do not have sharp boundaries and interiors or whose boundaries and interiors cannot be precisely determined.

The “Background” section discusses related work. The section “VASA: An Algebra for Vague Spatial Data in Databases” presents the Vague Spatial Algebra, and the section “FUSA: An Algebra for Fuzzy Spatial Data in Databases” introduces the Fuzzy Spatial Algebra. We provide a formal definition of the structure and semantics of both type systems. Further, we introduce spatial set operations for them and thus obtain vague and fuzzy versions of the geometric intersection, union, and difference operations. Finally, the section “Querying with VASA and FUSA” describes how these data types can be embedded into extensible databases and their query languages, and shows some example queries. This chapter concludes with some future research directions and conclusions.

BACKGROUND

We discuss related work that is relevant for the design and construction of VASA and FUSA. First we deal with crisp, determinate spatial data types and operations that serve as the basis of both algebras. Then we characterize spatial vagueness, give a classification of current models for representing spatial indeterminacy, and provide an overview of other models that are related to both algebras.

Crisp Spatial Data Types and Operations

In the spatial database and GIS community, *spatial data types* like *point*, *line*, or *region* have found wide acceptance as fundamental abstractions for modeling the structure of geometric entities, their

relationships, properties, and operations. They form the basis of a large number of data models and query languages for processing spatial data and have gained access into commercial software products. The literature distinguishes *simple* spatial data types (for example, Egenhofer 1994, Güting 1988) and *complex* spatial data types (for example, Clementini & Di Felice 1996a, Schneider 1997, Schneider & Behr 2006), depending on the spatial complexity they are able to model. Simple spatial data types (Figure 1(a)-(c)) only provide simple object structures like single points, continuous lines, and simple regions. However, from an application perspective, they are insufficient to cope with the variety and complexity of geographic reality. From a formal perspective, they are not closed under the geometric set operations *intersection*, *union*, and *difference*. Complex spatial data types (Figure 1(d)-(f)) solve these problems. They provide universal and versatile spatial objects with multiple components, permit regions with holes, are closed under geometric set operations, and form the basis of our vague spatial data types.

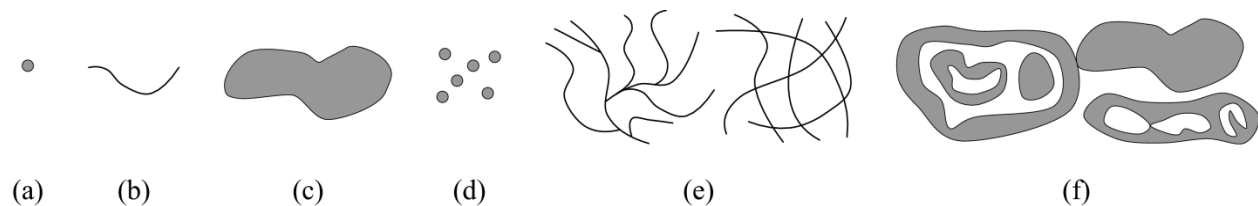


Figure 1. Examples of a simple point object (a), a simple line object (b), a simple region object (c), a complex point object (d), a complex line object (e), and a complex region object (f).

A Classification of Models for Vague Spatial Objects

So far, spatial data modeling has represented spatial objects as entities with sharply determined boundaries emphasizing abrupt changes of spatial phenomena. Hence, in the past, there has been a tendency to force geographical reality into determinate spatial objects. In practice, however, there is no apparent reason for the whole contour of a line or the boundary or interior of a region to be determinate. Numerous geographical application examples illustrate that the extent and the boundaries of spatial objects can be indeterminate. For instance, boundaries of geological, soil, and vegetation units are often crisp in some places and vague in others; concepts like the “Indian Ocean”, “South England”, or a biotope are intrinsically vague.

In the real, non-artifactual world, we can essentially find two categories of indeterminate boundaries: sharp boundaries whose position and shape are unknown or cannot be measured precisely, and boundaries which are not well defined or which are useless (for example, between a mountain and a valley). Spatial objects with indeterminate boundaries are difficult to represent and are so far not supported in spatial database systems and GIS. *Spatial vagueness* describes the feature of a spatial object that we cannot be sure whether certain components belong completely or partially to the object or not. According to the two categories of boundaries, mainly two kinds of spatial vagueness can be identified: spatial uncertainty and spatial fuzziness. *Spatial uncertainty* is traditionally equated with randomness and chance occurrence and relates either to a lack of knowledge about the position and shape of a spatial object with an existing, real boundary (*positional uncertainty*) or to the inability of measuring such an object precisely (*measurement uncertainty*). *Spatial fuzziness* is an intrinsic feature of a spatial object itself and describes the vagueness of such an object which certainly has an extent but which inherently cannot or does not have a precisely definable boundary.

At least four alternatives are proposed as general design methods: exact object models, fuzzy set-based models, probabilistic models, and rough set-based models.

Exact object models transfer data models, type systems, and concepts for spatial objects with sharp boundaries to objects with indeterminate boundaries. A benefit of this approach is that existing

definitions, techniques, data structures, algorithms, etc. need not be redeveloped but only modified and extended, or simply used. The approaches in Clementini & Di Felice (1996b), Clementini & Di Felice (2001), Cohn & Gotts (1996), and Schneider (1996) propose a *zone concept*. The central idea is to consider determined zones surrounding the indeterminate boundaries of a region (*broad boundaries*) and expressing its minimal and maximal extent. The zones serve as a description and separation of the space that certainly belongs to the region, that perhaps belongs to the region, and that is certainly outside. The approaches in Erwig & Schneider (1997) and Pauly & Schneider (2004), which contribute to the VASA approach, generalize these concepts and introduce *vague spatial data types* for *vague points*, *vague lines*, and *vague regions*. A vague spatial object is modeled as a pair of two meeting or disjoint crisp spatial objects. For example, a vague region is given as a pair of two meeting or disjoint crisp regions. The first object is the *kernel part* and models the part that definitely belongs to the vague spatial object. The second object is the *conjecture part* from which we can only assume that it or parts of it belong to the vague object. Designs of vague topological predicates, which are not discussed in this chapter, have been described in Clementini & Di Felice (1996b), Clementini & Di Felice (2001), Pauly & Schneider (2005a), Pauly & Schneider (2005b), and Pauly & Schneider (2006).

Fuzzy set-based models are based on *fuzzy set theory* (Zadeh 1965). It describes the *admission of the possibility* (given by a so-called *membership function*) that an individual is a member of a set or that a given statement is true. Hence, the vagueness represented by fuzziness is not the uncertainty of expectation. It is the vagueness resulting from the imprecision of meaning of a concept. Examples of fuzzy spatial objects include mountains, valleys, biotopes, and oceans, which cannot be rigorously bounded by a sharp line. The usefulness of fuzzy concepts in geoscience applications from a modeling standpoint has, for example, been demonstrated in Burrough, van Gaans & Macmillan (2000) for modeling geomorphological units, De Gruijter, Walvoort & Vangaans (1997), Lagacherie, Andrieux & Bouzigues (1996) for representing soil types and boundaries, Cheng, Molenaar & Lin (2001) for designing landscape objects, Brown (1998) for describing forest types, Hendricks Franssen, van Eijnsbergen & Stein (1997) for determining soil pollution classes in environmental applications, and Bogárdi, Bárdossy & Duckstein (1990) for performing hydrological studies. Approaches nearer to computer science have been given by Burrough (1996) introducing fuzzy geographical objects for modeling natural objects with indeterminate boundaries, Dutta (1989, 1991) and Kollias & Voliotis (1991) dealing with qualitative spatial and temporal reasoning using fuzzy logic, Edwards (1994) and Wang & Hall (1996) dealing with fuzzy representations of geographical boundaries in GIS, Usery (1996) defining concepts like core and boundary of a fuzzy region, and Wang, Hall & Subaryono (1990) and Wang (1994) presenting a fuzzy query approach in order to introduce more natural language expressions into GIS user interfaces. Approaches that deal with the issue of representing, storing, retrieving, and querying fuzzy spatial objects in *fuzzy spatial databases* are rare. However, there is an increasing interest in applying fuzzy theory to spatial topics. Examples are given by Altman (1994) presenting fuzzy set theoretic approaches for handling imprecision in spatial analysis and introducing *fuzzy regions* in \mathbb{N}^2 , Schneider (1999) providing the foundation of FUSA in terms of *fuzzy spatial data types* for *fuzzy points*, *fuzzy lines*, and *fuzzy regions* together with some *fuzzy spatial operations*, Dilo, de By & Stein (2007) introducing a similar type system of fuzzy (called “vague”) spatial data types with a more comprehensive set of operations, Schneider (2003) proposing a conceptual model and an implementation model of fuzzy spatial objects that are not defined on the Euclidean plane but on a discrete geometric domain called *grid partition*, and Schneider (2000) presenting metric operations on fuzzy spatial objects like the area of a fuzzy region or the length of a fuzzy line. Different models of fuzzy topological predicates, which are not discussed in this chapter and which characterize the relative position of two fuzzy spatial objects towards each other, are discussed in Petry et al. (2002), Schneider (2001a, 2001b), Shi & Guo (1999), Tang & Kainz (2002), and Zhan (1997, 1998).

Probabilistic models are able to represent positional and measurement uncertainty. Their basis is *probability theory* which defines the grade of membership of an entity in a set by a statistically defined probability function and deals with the expectation of a future event, based on something known now. As

examples of a large amount of literature, we mention the approaches in Blakemore (1984), Burrough (1996), Finn (1993), Shibasaki (1993), and Zinn, Bosch & Gertz (2007).

Rough set-based models are based on *rough set theory* (Pawlak 1982) and model vague spatial objects by a lower and an upper spatial approximation. As examples, Beaubouef, Ladner, & Petry (2004) and Worboys (1998) show how rough sets can be used for vague spatial data modeling.

VASA: AN ALGEBRA FOR VAGUE SPATIAL DATA IN DATABASES

VASA comprises a set of *vague spatial data types* designed as abstract data types and equipped with a set of operations between these types. A few operations are applicable to several types and are thus overloaded. Due to space restrictions, regarding operations, we confine ourselves to geometric set operations. The VASA approach belongs to the class of exact models. One of its main benefits is that its formal framework is based on well known and general crisp spatial data types. That is, vague spatial data types and operations are based on their crisp counterparts and can be expressed by them so that we obtain *executable specifications*. These specifications can be directly used as implementations and minimize the needed implementation effort since available implementations of crisp spatial algebras can be utilized. Moving from an exact to a vague domain does not necessarily invalidate conventional (computational) geometry; in our case, it is merely an extension. Thus, VASA provides the advantage of conceptual simplicity, ease of understanding, ease of use, ease of implementation, and efficiency. However, VASA is conservative and restrictive in the sense that it makes only use of a three-valued (and not a multi-valued) logic with the values *true*, *false*, and *maybe*. Therefore, VASA cannot capture all potential applications dealing with indeterminate spatial data. For example, fuzzy approaches allow the distinction and consideration of different degrees of spatial vagueness due to their multi-valued logic. However, this requires more and especially precise knowledge about uncertainty distributions that is frequently not available.

First, we describe and illustrate our concept of vague spatial objects. Then we formally define *vague spatial data types* for *vague points*, *vague lines*, and *vague regions*. Next we motivate vague spatial operations, and finally provide the formal definition of vague spatial set operations.

What are Vague Spatial Objects?

The central idea of *vague spatial objects* is to base their definition on already well defined, geometric modeling techniques. VASA leverages general exact object models incorporating the crisp spatial data types *point2D*, *line2D*, and *region2D*, as they have been reviewed before. We assume that these data types are closed under (appropriately defined) geometric union, intersection, and difference operations.

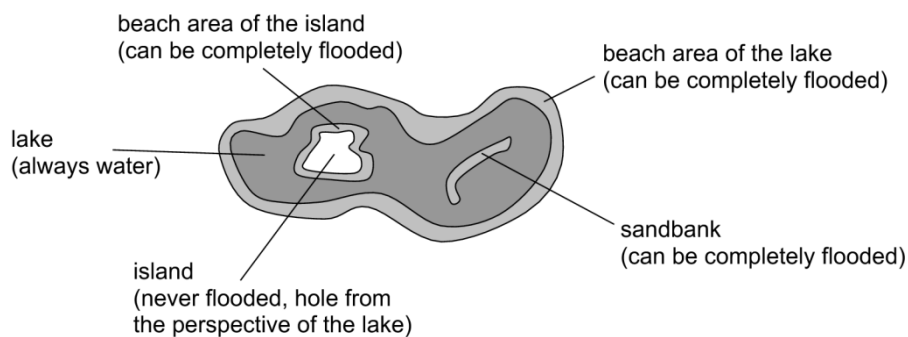


Figure 2. The extent of a lake depending on the degree of evaporation and on the amount of precipitation.

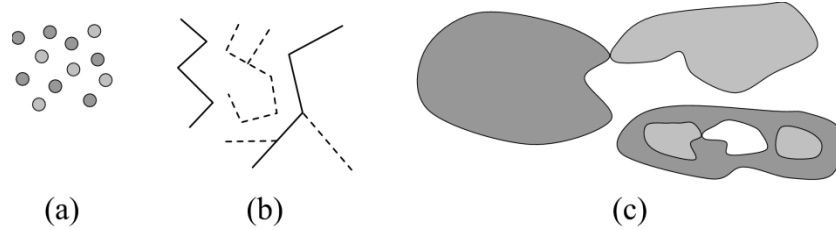


Figure 3. Examples of a (complex) vague point object (a), a (complex) vague line object (b), and a (complex) vague region object (c). The term “complex” indicates that each collection of components forms a single vague spatial object.

Vague spatial objects can, for example, represent the uncertainty about the precise paths or the spatial extent of phenomena in space; that is, objects can shrink and extend and hence have a minimal and maximal extent. An example is a lake whose water level depends on the degree of evaporation and on the amount of precipitation. High evaporation implies dry periods and thus a minimal water level. High precipitation entails rainy periods and thus a maximal water level. Islands in the lake are less flooded by water in dry periods and more flooded in rainy periods. If an island can never be completely flooded by water, it forms a “hole” in the lake. But if an island like a sandbank can be flooded completely, it belongs to the vague part of the lake. Hence, we have confident information about the minimal and maximal extent of a lake. But the actual extent of a lake, which is somewhere between these two extreme limits, is vague. Figure 2 illustrates this spatial constellation. Dark-gray shading shows areas that definitely belong to the lake. Light-gray shading indicates areas that perhaps belong to the lake. White color indicates areas that do definitely not belong to the lake.

As a further illustrating example, which we also use to introduce our terminology deployed in VASA, we consider a homeland security scenario. Secret services (should) have knowledge of the whereabouts of terrorists. For each terrorist, some of their refuges are precisely known; some others are assumed and thus only conjectures. We can model all these locations as a single *vague point* object where the precisely known locations are called the *kernel point* object and the assumed locations are denoted as the *conjecture point* object. Secret services are also interested in the routes a terrorist takes to move from one refuge to another. These routes can be modeled as a single *vague line* object. Some routes collected in a *kernel line* object have been identified with certainty. Other routes can only be assumed to be taken by a terrorist; they are gathered in a *conjecture line* object. Knowledge about areas of terroristic activities is also important for secret services. From some areas it is well known that a terrorist operates in them; we summarize them in a *kernel region* object. From other areas we can only assume that they are the target of terroristic activity; we denote them as a *conjecture region* object. Figure 3 gives some examples. Dark-gray shaded areas, straight curves, and dark-gray points indicate kernel parts. Areas with light-gray interiors, dashed lines, and light-gray points refer to conjecture parts. White areas describe exterior parts. In this sense, many application scenarios can be found that could leverage this concept of vague spatial objects.

A Generic Definition of Vague Spatial Data Types

Based on the motivation in the previous subsection, we now give formal definitions of *vague spatial data types*. An interesting observation is that these definitions can be given in a generic manner in the sense that type-specific considerations and distinctions are unnecessary. For the definition of vague points, vague lines, and vague regions we make use of the data types *point2D* for (complex) crisp points, *line2D* for (complex) crisp lines, and *region2D* for (complex) crisp regions, which are defined as special point sets (Schneider & Behr 2006) and closed under the geometric set operations \oplus (*union*), \otimes (*intersection*), \ominus (*difference*), and \sim (*complement*). Given a type $\alpha \in \{\textit{point2D}, \textit{line2D}, \textit{region2D}\}$, the signatures of the operations are $\oplus, \otimes, \ominus : \alpha \times \alpha \rightarrow \alpha$ and $\sim : \alpha \rightarrow \alpha$. Each type α together with the operations \oplus and \otimes

forms a Boolean algebra. We denote the identity of \otimes by $\mathbf{1}$, which corresponds to \mathbb{R}^2 . We represent the identity of \oplus by $\mathbf{0}$, which corresponds to the empty spatial object (empty point set \emptyset). The use of an exact model for constructing vague spatial data types leads to the benefit that existing definitions, techniques, data structures, and algorithms need not be redeveloped but can simply be used or in the worst case slightly modified or extended as necessary. This leads to the following generic definition of vague spatial data types:

Definition 1 Let $\alpha \in \{\text{point2D}, \text{line2D}, \text{region2D}\}$, and let A° denote the interior of a spatial object A . A *vague spatial data type* is given by a type constructor v as a pair of equal crisp spatial data types α , that is,

$$v(\alpha) = \alpha \times \alpha$$

such that for $w = (w_k, w_c) \in v(\alpha)$ and an auxiliary function $\text{points} : v(\alpha) \rightarrow 2^{\mathbb{R}^2}$, which yields the (unknown) point set of a vague spatial object, holds:

- (i) $w_k^\circ \cap w_c^\circ = \emptyset$
- (ii) $w_k \subseteq \text{points}(w) \subseteq w_k \oplus w_c$
- (iii) $\text{points}(w) \in \alpha$

We call $w \in v(\alpha)$ a (two-dimensional) vague spatial object with kernel part w_k and conjecture part w_c . Further, we call $w_o := \sim(w_k \oplus w_c)$ the *outside part* of w . If $\alpha = \text{point2D}$ holds, an element of $v(\text{point2D}) = \text{point2D} \times \text{point2D} =: v\text{point2D}$ is called a *vague point* object. Correspondingly, an element of $v(\text{line2D}) =: v\text{line2D}$ is called a *vague line* object, and an element of $v(\text{region2D}) =: v\text{region2D}$ is called a *vague region* object. If $w_k = \mathbf{0}$ and $w_c = \mathbf{0}$, we call $w = (\mathbf{0}, \mathbf{0})$ the *empty vague spatial object*.

Syntactically, a vague spatial object is described as a pair of crisp complex spatial objects of the same type. Semantically, four constraints have to be satisfied. In condition (i) we require that the kernel part and the conjecture part have disjoint interiors since a point of the same object cannot belong to both parts.

Condition (ii) makes two statements. First, it states that the kernel part w_k describes the determinate component of the vague spatial object, that is, the component that definitely and always belongs to the vague spatial object. Second, it states that the conjecture part w_c describes the vague component of the vague spatial object, that is, the component from which we cannot say with any certainty whether it or pieces of it belong to the vague spatial object or not. *Maybe* the conjecture part or pieces of it belong to the vague object; *maybe* this is not the case. We could also say that this is *unknown*. Another, related view is to regard w_k as a *lower* (minimal, guaranteed) *approximation* of w and $w_k \oplus w_c$ as an *upper* (maximally possible, speculative) *approximation* of w . This brings us near to a rough set theoretical view.

Condition (iii) requires that even if we do not know the exact point set of w , the actual point set $\text{points}(w)$ may not be arbitrary but must be compatible to type α . A direct conclusion from this requirement is that $\text{points}(w) \ominus w_k \in \alpha$ since $w_k \in \alpha$.

Finally, we give the definition of the *characteristic function* of a vague spatial object. This function decides about existence or non-existence of a point in a vague spatial object.

Definition 2 The *characteristic function* $\chi_w(p)$ for a point $p \in \mathbb{R}^2$ and a vague spatial object $w \in v(\alpha)$ is defined as:

$$\chi_w(p) = \begin{cases} \{1\} & \text{if } p \in w_k \\ \{0\} & \text{if } p \in \mathbb{R}^2 - (w_k \cup w_c) \\ \{0, 1\} & \text{if } p \in w_c - w_k \end{cases}$$

Note the deliberate use of set-theoretic operations. Especially the common boundary points of w_k and w_c ($w_k \cup w_c \neq \emptyset$) are mapped to 1. Further, we obtain that $\chi_w(p) = 1$ for all $p \in \text{points}(w)$.

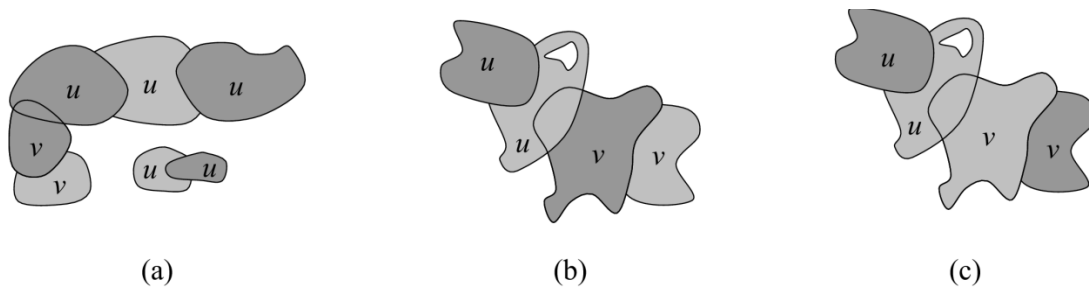


Figure 4. Example scenarios with a focus on the intersection between two kernel parts (a), a kernel part and a conjecture part (b), and two conjecture parts (c) of two vague regions.

What are Vague Spatial Operations?

Unsurprisingly, *vague spatial operations* are spatial operations operating on vague spatial objects. That is, these operations take vague spatial objects as arguments and yield vague objects as a result. It is common consensus in the spatial database field that a spatial algebra is never complete since always new operations can be designed and added to it. Hence, we take an extensible approach and assume that VASA will be extended if necessary.

We now briefly present two real life applications and motivate the use of vague spatial operations. The first example is taken from the animal kingdom. We view the living spaces of different animal species and distinguish kernel parts where they mainly live and conjecture parts like peripheral areas or corridors where they in particular hunt for food or which they cross in order to migrate from one kernel part to another one. We consider the following example queries regarding their living spaces: (1) Find the animal species that (partially) share their living spaces. (2) Determine hunters that penetrate into the living space of other animals. (3) Ascertain the areas where two species can only meet by accident. For two animal species u and v , the interesting situations for the queries are illustrated in Figure 4. The common task of all three queries is to compute the common living spaces of two animal species, that is, to calculate the intersection of two vague regions. But the nature of the intersection is different in all three cases. The first query asks for an intersection between two kernel parts (Figure 4(a)). The (non-empty) result is definitely a kernel part. The second query amounts to an intersection between a kernel part and a conjecture part but not between two kernel parts (Figure 4(b)). Since a conjecture part is involved that describes a vague part, we cannot make a definite statement whether this intersection belongs to the kernel part. It only remains to regard this intersection as a conjecture part. The third query exclusively asks for an intersection between two conjecture parts (Figure 4(c)); it is definitely a conjecture part. In total, we see that the “strength of intersection” decreases from left to right in Figure 4. Last but not least, we mention that the intersection of an exterior part with anything else is an exterior part.

The second example refers to our homeland security scenario from before. Taking into account spatial vagueness, we are able to pose interesting queries. We can ask for the locations where any two terrorists have definitely, perhaps, and/or never taken the same refuge. We can determine those terrorists that definitely, perhaps, and/or never operated in the same area. We can compute the locations where routes taken by different terrorists definitely, perhaps, and/or never crossed each other. We can find out the sphere of a number of terrorists on the basis of their locations as a *vague convex hull*.

Many further application scenarios and queries are conceivable. Vague concepts offer a greater flexibility and more nuances for modeling and computing properties of spatial phenomena in the real world than determinate “black or white” concepts do. Still, vague concepts comprise the modeling power of determinate concepts as a special case.

<i>vunion</i>		<i>vintersection</i>		<i>vdifference</i>		<i>vcomplement</i>	
●	● ● ○	●	● ● ○	●	● ● ○	● ● ○	
●	● ● ●	●	● ● ○	●	○ ● ●	○ ● ●	
●	● ● ●	●	● ● ○	●	○ ● ●	○ ● ●	
○	● ● ○	○	○ ○ ○	○	○ ○ ○	○ ● ●	
	(a)		(b)		(c)		(d)

Table 1. Components resulting from intersecting kernel parts, conjecture parts, and outside parts of two vague spatial objects with each other for the four vague geometric set operations *vunion* (a), *vintersection* (b), *vdifference* (c), and *vcomplement* (d).

A Generic Definition of Vague Spatial Operations

Due to space limitations, we confine our consideration of vague spatial operations to the vague geometric set operations *vunion*, *vintersection*, and *vdifference* that have all the same signature $v(\alpha) \times v(\alpha) \rightarrow v(\alpha)$ with $\alpha \in \{\text{point, line, region}\}$ in a type-independent and thus generic manner. In addition, we define the operation *vcomplement* with the signature $v(\alpha) \rightarrow v(\alpha)$. In order to define them for two vague spatial objects u and w , it is helpful to consider meaningful relationships between the kernel part, the conjecture part, and the outside part of u and w (Table 1). For each operation we give a table where a column/row labeled by ●, ●, or ○ denotes the kernel part, conjecture part, or outside part of u and w respectively. Each entry of the table denotes a possible combination, i.e., non-empty intersection, of kernel parts, conjecture parts, and outside parts of both objects, and the label in each entry specifies whether the corresponding intersection belongs to the kernel part, conjecture part, or outside part of the operation's result object.

Regarding the *union* operation (Table 1(a)), the intersection component of a kernel part with any other part belongs to the kernel part of the resulting vague spatial object r since the union operation asks for the definite membership in either part only. Likewise, the intersection component of the two conjecture parts or the intersection component of a conjecture part with an outside part belongs to the conjecture part of r , and only the intersection component of the two outside parts belongs to the outside part of r . Regarding the *intersection* operation (Table 1(b)), the intersection component of an outside part with any other part belongs to the outside part of r because intersection requires at least potential membership in both parts. The kernel part of r only contains components which definitely belong to the kernel parts of both operand objects. The intersection component of both conjecture parts or of a conjecture part and a kernel part contribute to the conjecture part of r . Obviously, the *complement* (Table 1(d)) of the kernel part is the outside part, and vice versa. With respect to the conjecture part, anything inside the vague part of an object might or might not belong to the object; hence, the result is the conjecture part itself. The *difference* (Table 1(c)) between any two parts is equal to the intersection of the first part with the complement of the second part.

Table 1 enables us now to formally define the four operations. An interesting aspect is that these definitions can be based solely on already known crisp geometric set operations on well-understood exact spatial objects. Hence, we are able to give *executable specifications* for the vague geometric set operations. That is, by employing the geometric set operations of an available crisp spatial algebra, we can directly *execute* the vague geometric set operations without being forced to design and implement new algorithms for them.

Definition 3 Let $u, w \in v(\alpha)$, and let u_k and w_k denote their kernel parts and u_c and w_c their conjecture parts. We define:

$$\begin{aligned}
\text{(i) } u \text{ vunion } w & := (u_k \oplus w_k, (u_c \oplus w_c) \ominus (u_k \oplus w_k)) \\
\text{(ii) } u \text{ vintersection } w & := (u_k \otimes w_k, (u_c \otimes w_c) \oplus (u_k \otimes w_k) \oplus (u_c \otimes w_k)) \\
\text{(iii) } u \text{ vdifference } w & := (u_k \otimes (\sim(w_k \oplus w_c)), (u_c \otimes w_c) \oplus (u_k \otimes w_k) \oplus (u_c \otimes (\sim(w_k \oplus w_c)))) \\
\text{(iv) } v \text{ complement } u & := (\sim(u_k \oplus u_c), u_c)
\end{aligned}$$

If u and v are vague spatial objects with empty conjecture parts, that is, $u_c = \mathbf{0}$ and $w_c = \mathbf{0}$, the operations behave exactly like their crisp counterparts.

One can show (not done here) that the specifications of Definition 3 fit to the specifications in Table 1.

FUSA: AN ALGEBRA FOR FUZZY SPATIAL DATA IN DATABASES

FUSA comprises a set of *fuzzy spatial data types* designed as abstract data types and equipped with a set of operations between these types. The FUSA approach belongs to the class of fuzzy set-based models. The formal framework of the algebra design is based on *fuzzy set theory* (Zadeh 1965, Buckley & Eslami 2002) and *fuzzy topology* (Liu & Luo 1997). FUSA makes use of a multi-valued logic and has the great advantage of providing a fine-grained view and distinction of spatial indeterminacy. However, this requires much more and especially precise data and knowledge about uncertainty distributions.

First, we describe and illustrate our concept of fuzzy spatial objects. Then we formally define *fuzzy spatial data types* for *fuzzy points*, *fuzzy lines*, and *fuzzy regions*. Finally we provide the formal definition of fuzzy spatial set operations.

What are Fuzzy Spatial Objects?

Points are the simplest geometric abstraction. They are the atoms, that is, the elements, of the Euclidean space \mathbb{R}^2 . Each element of a crisp point set X of \mathbb{R}^2 belongs definitely and totally to this set. For each point of \mathbb{R}^2 , we can definitely say whether it belongs to X or not. A crisp point object, as it has been defined, for example, in Schneider 1997 and Schneider & Behr 2006, includes a finite set of points in order to ensure the closure of set operations. In the fuzzy domain, a fuzzy point set \tilde{X} also represents a subset of \mathbb{R}^2 but each point of \tilde{X} only has a certain partial membership in this set. It can also partially belong to another fuzzy set, that is, multiple set membership is possible. For example, if we consider an air-polluted area (which is a fuzzy region), then each point of this area is fuzzy since the concentration of air pollution at this point is not 100 percent. That is, the degree of membership of this fuzzy point in the air-polluted area is less than 1 and larger than 0. Fuzzy points can also arise as the result of two fuzzy lines.

Lines are the one-dimensional, geometric abstraction for linear features like rivers, boundaries, and transportation routes. Each crisp line is a subset of the Euclidean plane \mathbb{R}^2 with particular properties. Each element of a single crisp line is a crisp point that definitely and totally belongs to the line. A crisp line object (Schneider 1997; Schneider & Behr 2006) includes a finite set of blocks. Each block consists of a finite set of simple lines (curves) such that each pair of simple lines is either disjoint or meets in a common end point. Fuzzy lines are supposed to adopt the fundamental structure of crisp lines. That is, a fuzzy line \tilde{X} has the same linear geometry as a crisp line and is hence a subset of \mathbb{R}^2 ; however, each point of \tilde{X} may only be to some degree a member of \tilde{X} . For example, the pollution of a river can be represented by the line geometry of the river where each point represents the degree or concentration of pollution at that location. The concentration is larger than 0 for all points of the fuzzy line but usually different at different locations.

A crisp region is defined as a set of disjoint, connected areal components, called *faces*, possibly with disjoint *holes* (Clementini & Di Felice 1996a, Schneider 1997, Schneider & Behr 2006) in the Euclidean space \mathbb{R}^2 . This model has the nice property that it is closed under (appropriately defined) geometric

union, intersection, and difference operations. Hence, we would like to transfer the structural definition of crisp regions (that is, the component view) to fuzzy regions and allow partial membership of a point in a region. There are at least three possible, related interpretations for a point in a fuzzy region. First, this situation may be interpreted as the *degree of belonging* to which that point is *inside* or *part of* some areal feature. Consider the transition between a mountain and a valley and the problem to decide which points have to be assigned to the valley and which points to the mountain. Obviously, there is no strict boundary between them, and it seems to be more appropriate to model the transition by partial and multiple membership. Second, this situation may indicate the *degree of compatibility* of the individual point with the attribute or concept represented by the fuzzy region. An example are “warm areas” where we must decide for each point whether and to which grade it corresponds to the concept “warm”. Third, this situation may be viewed as the *degree of concentration* of some attribute associated with the fuzzy region at the particular point.

An example is air pollution where we can assume the highest concentration at power stations, for instance, and lower concentrations with increasing distance from them. The distribution of attribute values within a region and transitions between different regions may be *smooth* or *continuous*. Application examples are air pollution, temperature zones, magnetic fields, storm intensity, and sun insolation. Figure 5 demonstrates a possible visualization of a fuzzy region object which could model the expansion of air pollution caused by two nearby power stations. The left image shows a radial expansion of the first power station where the degree of pollution concentrates in the center (darker locations) and decreases with increasing distance from the power station (brighter locations). The right image shows the distribution of air pollution of the second power station that is surrounded by high mountains to the north, the south, and the west. Hence, the pollution cannot escape in these directions and finds its way out of the valley in eastern direction.



Figure 5. Example of a fuzzy region consisting of two fuzzy faces as components.

A Formal Definition of Fuzzy Spatial Data Types

We now provide formal definitions of data types for fuzzy points, fuzzy lines, and fuzzy regions. We assume that the reader is familiar with the fundamentals of fuzzy set theory and fuzzy topology. We begin with a definition of a single fuzzy point that can be used as a foundation for a definition of the type *fpoint*. This definition views a fuzzy point as a point in the two-dimensional Euclidean space with a membership value greater than 0, since 0 documents the non-existence of a point.

Definition 4 A *fuzzy point* \tilde{p} at (a, b) in \mathbb{IR}^2 , written $\tilde{p}(a, b)$, is a fuzzy singleton in \mathbb{IR}^2 defined by

$$\mu_{\tilde{p}(a, b)}(x, y) = \begin{cases} m & \text{if } (x, y) = (a, b) \\ 0 & \text{otherwise} \end{cases}$$

with $0 < m \leq 1$. Point \tilde{p} is said to have *support* $\text{supp}(\tilde{p}(a, b)) = \{(a, b)\}$ and value m . Let P_f be the set of all fuzzy points. P_f is a proper superset of P_c , the set of all crisp points in \mathbb{IR}^2 . For $\tilde{p} = p = (a, b) \in P_c$, we

obtain $\mu_{p-(a,b)}(x,y) = \chi_p(x,y) = 1$, if $(x,y) = (a,b)$, and 0 otherwise.

Next, we define three *geometric primitives* on fuzzy points.

Definition 5 Let $\tilde{p}(a,b), \tilde{q}(c,d) \in P_f$ with $a,b,c,d \in \mathbb{R}$. Then

- (i) $\tilde{p}(a,b) = \tilde{q}(c,d) \quad :\Leftrightarrow \quad a = c \wedge b = d \wedge \mu_{p-(a,b)} = \mu_{q-(a,b)}$
- (ii) $\tilde{p}(a,b) \neq \tilde{q}(c,d) \quad :\Leftrightarrow \quad \neg(\tilde{p}(a,b) = \tilde{q}(c,d))$
- (iii) $\tilde{p}(a,b)$ and $\tilde{q}(c,d)$ are *disjoint* $\quad :\Leftrightarrow \quad \text{supp}(\tilde{p}(a,b)) \cap \text{supp}(\tilde{q}(c,d)) = \emptyset$

In contrast to crisp points, for fuzzy points we also have a predicate for disjointedness.

Definition 6 The fuzzy spatial data type *fpoint* is defined as

$$fpoint = \{Q \subseteq P_f \mid \forall \tilde{p}, \tilde{q} \in Q : \tilde{p}(a,b) \text{ and } \tilde{q}(c,d) \text{ are disjoint} \wedge Q \text{ is finite}\}$$

Disjointedness of the single fuzzy points of a fuzzy point object is required since the membership degree of each single fuzzy point should be unique.

Next, we specify the fuzzy spatial data type *fline* for *fuzzy lines*. For that, we first introduce a *simple* fuzzy line as a continuous curve with smooth transitions of membership grades between neighboring points of the line (Figure 6(a)). We assume a total order on \mathbb{R}^2 which is given by the lexicographic order “<” on the coordinates (first x , then y) of the points of \mathbb{R}^2 .

Definition 7 The membership function of a *simple fuzzy line* \tilde{l} is defined by

$\mu_{\tilde{l}}: f_{\tilde{l}} \rightarrow [0, 1]$ with $f_{\tilde{l}}: [0, 1] \rightarrow \mathbb{R}^2$ such that

- (i) $\mu_{\tilde{l}}$ is continuous
- (ii) $f_{\tilde{l}}$ is continuous
- (iii) $\forall a, b \in]0, 1[: a \neq b \Rightarrow f_{\tilde{l}}(a) \neq f_{\tilde{l}}(b)$
- (iv) $\forall a \in \{0, 1\} \forall b \in]0, 1[: f_{\tilde{l}}(a) \neq f_{\tilde{l}}(b)$
- (v) $f_{\tilde{l}}(0) < f_{\tilde{l}}(1) \vee (f_{\tilde{l}}(0) = f_{\tilde{l}}(1) \wedge \forall a \in]0, 1[: f_{\tilde{l}}(0) < f_{\tilde{l}}(a))$

Function $f_{\tilde{l}}$ on its own models a continuous, simple *crisp* line (a *curve*). The points $f_{\tilde{l}}(0)$ and $f_{\tilde{l}}(1)$ are called the *end points* of f . The definition allows loops ($f_{\tilde{l}}(0) = f_{\tilde{l}}(1)$) but prohibits equality of interior points and thus self-intersections (condition (iii)). The reason is that self-intersections do not occur in spatial reality; hence, our model excludes them. Condition (iv) disallows the equality of an interior point with an end point. Condition (v) requires that in a *closed simple line* $f_{\tilde{l}}(0)$ must be the leftmost point, that is, the smallest point with respect to the lexicographic order “<”. The main reason of the conditions (iii) to (v) is to ensure the uniqueness of representation of a simple fuzzy line. For example, consider a simple line that intersects itself one time and is not closed. An alternative to model this configuration is to model the intersection point explicitly with three incident simple lines from which one line forms a closed loop. This alternative concept leads to a unique representation.

All conditions together define a fuzzy line \tilde{l} as the fuzzy point set $\tilde{l} = \{(p, \mu_{\tilde{l}}(p)) \mid p \in f_{\tilde{l}}([0, 1])\}$.

Definition 8 Let SL_f be the set of fuzzy simple lines, and let $\tilde{l}_1, \tilde{l}_2 \in SL_f$. Then we define:

- (i) \tilde{l}_1, \tilde{l}_2 are *disjoint* $\quad :\Leftrightarrow \quad \text{supp}(\tilde{l}_1) \cap \text{supp}(\tilde{l}_2) = \emptyset$
- (ii) \tilde{l}_1, \tilde{l}_2 *meet* $\quad :\Leftrightarrow \quad f_{\tilde{l}_1}(]0, 1[) \cap f_{\tilde{l}_2}(]0, 1[) = \emptyset \wedge \{f_{\tilde{l}_1}(0), f_{\tilde{l}_1}(1)\} \cap \{f_{\tilde{l}_2}(0), f_{\tilde{l}_2}(1)\} \neq \emptyset$



Figure 6. Example of a simple fuzzy line (a) and a (complex) fuzzy line (b). Fuzziness is indicated by shading. The complex fuzzy line consists of two fuzzy blocks that are made up of seven fuzzy simple lines.

Definition 9 A fuzzy block \tilde{b} is the fuzzy set $\tilde{b} = \cup_{i=1}^n \tilde{l}_i$ such that¹

- (i) $n \in \mathbb{N}, 1 \leq i \leq n : \tilde{l}_i \in SL_f$
- (ii) $\forall 1 \leq i < j \leq n : f_{\tilde{l}_i}([0, 1]) \cap f_{\tilde{l}_j}([0, 1]) = \emptyset$
- (iii) $\forall 1 \leq i < j \leq n : \{f_{\tilde{l}_i}(0), f_{\tilde{l}_i}(1)\} \cap f_{\tilde{l}_j}([0, 1]) = \emptyset \wedge f_{\tilde{l}_i}([0, 1]) \cap \{f_{\tilde{l}_j}(0), f_{\tilde{l}_j}(1)\} = \emptyset$
- (iv) $\forall 1 \leq i \leq n \exists 1 \leq j \leq n, j \neq i : f_{\tilde{l}_i}(0), f_{\tilde{l}_i}(1) \cap \{f_{\tilde{l}_j}(0), f_{\tilde{l}_j}(1)\} \neq \emptyset$
- (v) $\forall 1 \leq i, j \leq n \forall a, k \in \{0, 1\}$ let $V_{\tilde{l}_i}^a = \{(j, k) \mid f_{\tilde{l}_i}(a) = f_{\tilde{l}_j}(k)\}$. Then we require:
 $\forall 1 \leq i \leq n \forall a \in \{0, 1\} : (|V_{\tilde{l}_i}^a| = 1) \vee (|V_{\tilde{l}_i}^a| > 2)$
- (vi) $\forall 1 \leq i \leq n \forall a \in \{0, 1\} \forall (j, k) \in V_{\tilde{l}_i}^a : \mu_{\tilde{l}_i}(f_{\tilde{l}_i}(a)) = \mu_{\tilde{l}_j}(f_{\tilde{l}_j}(k))$

Intuitively, a fuzzy block is a maximal, connected fuzzy line component (Figure 6(b)). Condition (i) states that a fuzzy block consists of a finite set of fuzzy simple lines. Condition (ii) requires that the elements of a fuzzy block do not intersect or overlap within their interior. Moreover, they may not be touched within their interior by an endpoint of another element (condition (iii)). The main reason for both conditions is again their uniqueness of representation. Condition (iv) ensures the property of connectivity of a fuzzy block; isolated fuzzy simple lines are disallowed. Condition (v) expresses that each endpoint of an element of \tilde{b} must belong to exactly one or more than two incident elements of \tilde{b} (note that always $(i, a) \in V_{\tilde{l}_i}^a$). This condition supports the requirement of “maximal elements” and hence achieves uniqueness of representation. Condition (vi) requires that the membership values of more than two elements of \tilde{b} with a common end point must have the same membership value; otherwise we get a contradiction saying that a point of a fuzzy block has more than one different membership value.

Definition 10 Let B_f be the set of fuzzy blocks over SL_f . The disjointedness of any two fuzzy blocks $\tilde{b}_1, \tilde{b}_2 \in B_f$ is defined as follows:

$$\tilde{b}_1 \text{ and } \tilde{b}_2 \text{ are disjoint} \quad : \Leftrightarrow \quad \text{supp}(\tilde{b}_1) \cap \text{supp}(\tilde{b}_2) = \emptyset$$

We are now able to define the fuzzy spatial data type *fline* for fuzzy lines in Definition 11 and give an example of a fuzzy line object in Figure 6(b).

Definition 11 The fuzzy spatial data type *fline* is defined as

$$\text{fline} = \{\cup_{i=1}^n \tilde{b}_i \mid n \in \mathbb{N} \wedge \forall 1 \leq i \leq n : \tilde{b}_i \in B_f \wedge \forall 1 \leq i < j \leq n : \tilde{b}_i \text{ and } \tilde{b}_j \text{ are disjoint}\}$$

¹ The application of a function f to a set X of values is defined as $f(X) = \{f(x) \mid x \in X\}$.



Figure 7. Example of a fuzzy set that is not fuzzy region due to lower-dimensional, geometric anomalies like cuts, punctures, and dangling lines.

Finally, we specify the fuzzy spatial data type *fregion* for *fuzzy regions*. Since our objective is to model two-dimensional fuzzy areal objects for spatial applications, we consider a fuzzy topology \mathcal{T} on the Euclidean plane \mathbb{R}^2 . In this spatial context, we denote the elements of \mathcal{T} as *fuzzy point sets*. The membership function for a fuzzy point set \tilde{A} in the plane is then described by $\mu_{\tilde{A}} : \mathbb{R}^2 \rightarrow [0, 1]$. From an application point of view, there are two observations that prevent a definition of a fuzzy region simply as a fuzzy point set.

The first observation refers to a necessary *regularization* of fuzzy point sets. The first reason for this measure is that fuzzy (as well as crisp) regions that actually appear in spatial applications in most cases cannot be just modeled as arbitrary point sets but have to be represented as point sets that do not have “geometric anomalies” (Figure 7) and that are in a certain sense *regular*. Geometric anomalies relate to isolated or dangling line or point features and missing lines and points in the form of cuts and punctures. Spatial phenomena with such degeneracies never appear as entities in reality. The second reason is that, from a data type point of view, we are interested in fuzzy spatial data types that satisfy closure properties for (appropriately defined) geometric union, intersection, and difference.

We are, of course, confronted with the same problem in the crisp case where the problem can be avoided by the concept of *regularity* (Tilove 1980; Schneider 1997). It turns out to be useful to appropriately transfer this concept to the fuzzy case.

Definition 12 Let \tilde{A} be a fuzzy set of a fuzzy topological space $(\mathbb{R}^2, \mathcal{T})$. Then \tilde{A} is called a *regular open fuzzy set* if $\tilde{A} = \text{int}_{\mathcal{T}}(\text{cl}_{\mathcal{T}}(\tilde{A}))$.

Whereas crisp regions are usually modeled as *regular closed crisp sets*, we will use *regular open fuzzy sets* due to their vagueness and their usual lack of boundaries. Regular open fuzzy sets avoid the aforementioned geometric anomalies, too. Since application examples show that fuzzy regions can also be partially bounded, we admit *partial boundaries* with a crisp or fuzzy character and call them *frontiers*.

Definition 13 The *frontier* of a fuzzy set is defined as

$$\text{frontier}_{\mathcal{T}}(\tilde{A}) := \{(x, y), \mu_{\tilde{A}}(x, y) \mid (x, y) \in \text{supp}(\tilde{A}) - \text{supp}(\text{int}_{\mathcal{T}}(\tilde{A}))\}$$

The term $\text{supp}(\tilde{A}) - \text{supp}(\text{int}_{\mathcal{T}}(\tilde{A}))$ determines the crisp locations of all fuzzy points of \tilde{A} that are not interior points. However, these locations do not necessarily all belong to the boundary of \tilde{A} since \tilde{A} has not been constrained so far. This is done in the following definition.

Definition 14 A fuzzy set \tilde{A} is called a *spatially regular fuzzy set* if, and only if,

- (i) $\text{int}_{\mathcal{T}}(\tilde{A})$ is a regular open fuzzy set
- (ii) $\text{frontier}_{\mathcal{T}}(\tilde{A}) \subseteq \text{frontier}_{\mathcal{T}}(\text{cl}_{\mathcal{T}}(\text{int}_{\mathcal{T}}(\tilde{A})))$
- (iii) $\text{frontier}_{\mathcal{T}}(\tilde{A})$ is a partition of $n \in \mathbb{N}$ connected boundary parts (fuzzy sets)

Not every set \tilde{A} is a spatially regular fuzzy set. Therefore, condition (i) ensures that the interior of \tilde{A} is without any geometric anomalies. The other two conditions arrange for a correct (partial) boundary if it exists. Condition (ii) works as follows: On the right side of “ \subseteq ”, the set $\tilde{A}' = cl_{\mathcal{F}}(int_{\mathcal{F}}(\tilde{A}))$ is a *regular closed fuzzy set*, that is, the interior of \tilde{A} is complemented by its boundary without any geometric anomalies. Hence, the *frontier* $_{\mathcal{F}}$ operator applied to \tilde{A}' yields the boundary of \tilde{A}' . The condition now requires that the frontier of \tilde{A} is a subset of the frontier of \tilde{A}' and does not contain other fuzzy points. Condition (iii) states that the frontier of \tilde{A} has to consist of a finite number of connected pieces due to the “finite component assumption” explained before. Infinitely many boundary pieces cannot be represented in an implementation.

From the definition of *frontier* $_{\mathcal{F}}$, we can conclude that $frontier_{\mathcal{F}}(\tilde{A}) = \emptyset$ if \tilde{A} is regular open. We will base our definition of fuzzy regions on spatially regular fuzzy sets and define a *regularization* function reg_f which associates the interior of a fuzzy set \tilde{A} with its corresponding regular open fuzzy set and which restricts the partial boundary of \tilde{A} (if it exists at all) to a part of the boundary of the corresponding regular closed fuzzy set of \tilde{A} .

Definition 15 The *fuzzy regularization function* reg_f applied to a fuzzy set \tilde{A} is defined as

$$reg_f(\tilde{A}) = int_{\mathcal{F}}(cl_{\mathcal{F}}(\tilde{A})) \cup (frontier_{\mathcal{F}}(\tilde{A}) \cap frontier_{\mathcal{F}}(cl_{\mathcal{F}}(int_{\mathcal{F}}(\tilde{A}))))$$

The different components of the regularization process work as follows: the *interior* operator $int_{\mathcal{F}}$ eliminates dangling point and line features since their interior is empty. The *closure* operator $cl_{\mathcal{F}}$ removes cuts and punctures by appropriately adding points. Furthermore, the *closure* operator introduces a fuzzy boundary (similar to a crisp boundary in the ordinary point-set topological sense) separating the points of a closed set from its exterior. The operator *frontier* $_{\mathcal{F}}$ supports the restriction of the boundary.

The second observation is that, according to the application cases shown before, the mapping $\mu_{\tilde{A}}$ itself may not be arbitrary but must take into account the intrinsic smoothness of fuzzy regions. This property can be modeled by the well known mathematical concept of *continuity*. Here, we employ the concept of a piecewise continuous function for modeling the smooth membership distribution in a single fuzzy face. A function is *piecewise continuous* if it is made of a finite number of continuous pieces. Hence, it has only a finite number of *discontinuities* (*continuity gaps*), and its left and right limits are defined at each discontinuity. The only possible kinds of discontinuities for a piecewise continuous function are removable and step discontinuities. A *removable discontinuity* represent a hole in the function graph. It can be repaired by filling in a single point. A *step discontinuity* (also called *semi-continuity*) is a location in the function graph where the graph steps or jumps from one connected piece of the graph to another. Formally, it is a discontinuity for which the limits from the left and right both exist but are not equal to each other.

We are now able to give the definition of the fuzzy spatial data type *fregion* for fuzzy regions.

Definition 16 Let *SRFS* be the set of spatially regular fuzzy sets. The fuzzy spatial data type *fregion* is defined as

$$\begin{aligned} fregion = \{ \tilde{R} \in SRFS \mid & \text{(i) } \tilde{R} = \bigcup_{i=1}^n \tilde{R}_i, n \in \mathbb{N} \\ & \text{(ii) } \forall 1 \leq i \leq n : R_i \text{ is a connected component (fuzzy face)} \\ & \text{(iii) } \mu_{\tilde{R}} = \bigcup_{i=1}^n \mu_{\tilde{R}_i} \\ & \text{(iv) } \forall 1 \leq i \leq n : \mu_{\tilde{R}_i} \text{ is a piecewise continuous function} \} \end{aligned}$$

Since different connected components of a set are disjoint (except for single common boundary points perhaps), the fuzzy faces of a fuzzy region object are disjoint too.

Fuzzy Spatial Set Operations

The fuzzy spatial set operations *funion*, *fintersection*, and *fdifference* have all the same signature $\alpha \times \alpha \rightarrow \alpha$ with $\alpha \in \{fpoint, fline, fregion\}$. In addition, we define the operation *fcomplement* with the signature $\alpha \rightarrow \alpha$. The definition of these operations for $\alpha = fpoint$ corresponds to the ones on fuzzy sets.

Definition 17 Let $\tilde{A}, \tilde{B} \in fpoint$. Then

- (i) \tilde{A} *funion* $\tilde{B} \quad := \tilde{A} \cup \tilde{B} = \{(x, y), \mu_{\tilde{A} \cup \tilde{B}}((x, y)) \mid (x, y) \in \mathbb{R}^2$
 $\quad \quad \quad \wedge \mu_{\tilde{A} \cup \tilde{B}}((x, y)) = \max(\mu_{\tilde{A}}((x, y)), \mu_{\tilde{B}}((x, y)))\}$
- (ii) \tilde{A} *fintersection* $\tilde{B} \quad := \tilde{A} \cap \tilde{B} = \{(x, y), \mu_{\tilde{A} \cap \tilde{B}}((x, y)) \mid (x, y) \in \mathbb{R}^2$
 $\quad \quad \quad \wedge \mu_{\tilde{A} \cap \tilde{B}}((x, y)) = \min(\mu_{\tilde{A}}((x, y)), \mu_{\tilde{B}}((x, y)))\}$
- (iii) \tilde{A} *fdifference* $\tilde{B} \quad := \tilde{A} - \tilde{B} = \tilde{A} \cap \neg \tilde{B}$
- (iv) *fcomplement* $\tilde{A} \quad := \neg \tilde{A} = \{(x, y), \mu_{\neg \tilde{A}}((x, y)) \mid (x, y) \in \mathbb{R}^2$
 $\quad \quad \quad \wedge \mu_{\neg \tilde{A}}((x, y)) = 1 - \mu_{\tilde{A}}((x, y))\}$

These operations are defined in a similar manner for $\alpha = fline$. However, since the result of such an operation can consist of linear and punctual parts, the punctual parts have to be removed by a corresponding regularization function.

For $\alpha = fregion$, we first, informally and without proof, make some statements about set operations on regular open fuzzy sets. The intersection of two regular open fuzzy sets is regular open. The union, difference, and complement of two regular open fuzzy sets are not necessarily regular open since they can produce anomalies. Correspondingly, this also holds for spatially regular fuzzy sets. Hence, we introduce *regularized set operations* on spatially regular fuzzy sets that preserve regularity.

Definition 18 Let $\tilde{A}, \tilde{B} \in fregion$, and let $a \dot{-} b = a - b$ for $a \geq b$ and $a \dot{-} b = 0$ otherwise ($a, b \in \mathbb{R}_0^+$). Then we define the following *regularized set operations*:

- (i) $\tilde{A} \cup_r \tilde{B} \quad := reg_f(\tilde{A} \cup \tilde{B})$
- (ii) $\tilde{A} \cap_r \tilde{B} \quad := reg_f(\tilde{A} \cap \tilde{B})$
- (iii) $\tilde{A} -_r \tilde{B} \quad := reg_f(\{(x, y), \mu_{\tilde{A} -_r \tilde{B}}((x, y)) \mid (x, y) \in supp(\tilde{A})$
 $\quad \quad \quad \wedge \mu_{\tilde{A} -_r \tilde{B}}((x, y)) = \mu_{\tilde{A}}((x, y)) \dot{-} \mu_{\tilde{B}}((x, y))\})$
- (iv) $\neg_r \tilde{A} \quad := reg_f(\neg \tilde{A})$

Note that we have changed the meaning of difference, that is, $\tilde{A} -_r \tilde{B} \neq \tilde{A} \cap_r \neg \tilde{B}$, since the right side of the inequality is not meaningful in the spatial context. Regular open fuzzy sets, spatially regular fuzzy sets, and regularized set operations express a natural formalization of the desired closure properties of fuzzy geometric set operations.

Querying with VASA and FUSA

The design of the data types provided by VASA and FUSA as abstract data types whose internal structure is hidden from the user enables us to embed these types as attribute types in all known database models (for example, relational, object-oriented, object-relational, complex models). Information about objects of these types can only be obtained by high-level operations and predicates and not by direct access to object components. Assuming an available implementation, we look at two example queries. Note that the query capabilities are restricted since we have not presented the full algebras with all operations and especially predicates.

Our first scenario makes use of VASA and assumes an ecological database with the following relations:

```
weather(climate: string; area: vregion)  
soil(quality: string, area: vregion)
```

The relation *weather* has a column named *area* containing vague region values for various climatic conditions given by the column *climate*. The relation *soil* describes the soil quality for certain regions.

The query considered is supposed to find out all regions of bad ecological conditions, that is, all locations where a lack of water or a bad soil quality is a hindrance for cultivation.

```
select  vunion(dry area, bad soil) as bad region  
from    select vsum(area) as dry area from weather where climate = "dry";  
        select vsum(area) as bad soil from soil where quality = "bad"
```

In the *from* clause, the function *vsum* is a *spatial aggregation* function that here computes the geometric union of a collection of selected vague point objects of the column *area*. We obtain two temporary relations that contain the aggregated areas of dry climate and bad soil quality respectively. Each relation contains a single tuple with a single attribute value of type *vregion*. In the *select* clause, we compute the union of the two attribute values of both tuples and obtain a result relation with a single attribute *bad region* of type *vregion* and a single tuple.

Our second scenario makes use of FUSA and takes an example from the animal kingdom. Information about animals species and their habitats is assumed to be stored in the following relation:

```
animal(name: string, habitat: fregion)
```

The fact that the habitat of an animal is modeled as a fuzzy region indicates that the extent of such a habitat is not clearly known; it is fuzzy. We consider the query that asks about the common living space of each pair of animal species. For this purpose, we use the special fuzzy modifier *at all* that represents the existential modifier and checks whether a predicate like *overlaps* can be fulfilled to any extent.

```
select  A.name, B.name, fintersection(A.habitat; B.habitat)  
from    animals A, animals B  
where  A.habitat at all overlaps B.habitat
```

FUTURE RESEARCH DIRECTIONS

Future research will first have to focus on an extension and completion of the available models VASA for vague spatial objects and FUSA for fuzzy spatial objects. This refers to comprehensive collections of vague and fuzzy spatial operations and predicates. In terms of predicates, especially vague and fuzzy topological and directional relationships are of interest.

Second, an interesting question is how VASA and especially FUSA can be implemented. While we can implement VASA on available software implementations of crisp spatial data types, the implementation of FUSA seems to be more difficult since the membership values within a fuzzy spatial object can change. For example, if we consider a point in a fuzzy region and a small circle around it, it could be that all points in the circle have a different membership value. Since the number of these points is infinite, this leads to a collision with the finiteness of computers. Appropriate finite approximations have to be found that enable us to represent such objects.

Third, VASA and FUSA as type systems have to be integrated into extensible database systems to make their functionality available to the user.

Fourth, a vague spatial query language for VASA and a fuzzy spatial query language for FUSA have to be designed and embedded into database systems.

CONCLUSION

In this chapter, we have demonstrated how two different soft computing techniques with different expressiveness can be leveraged for spatial data handling in the context of spatial databases and Geographic Information Systems. Two type systems or algebras named Vague Spatial Algebra (VASA) and Fuzzy Spatial Algebra (FUSA) have been sketched whose types are modeled as abstract data types and can be integrated as attribute data types in any (extensible) database system. The focus of this chapter has been on the design of the data types. The description of spatial operations and predicates as well as implementation issues have been neglected due to space limitations. However, references given below will lead the reader to the respective literature.

REFERENCES

- Altman, D. (1994). Fuzzy Set Theoretic Approaches for Handling Imprecision in Spatial Analysis. *International Journal of Geographical Information Systems*, 8(3), 271-289.
- Beaubouef, T., Ladner, R., & Petry F. (2004). Rough Set Spatial Data Modeling for Data Mining. *International Journal of Geographical Information Science*, 19, 567-584.
- Blakemore, M. (1984). Generalization and Error in Spatial Databases. *Cartographica*, 21(2/3), 131-139.
- Brown, D.G. (1998). Mapping Historical Forest Types in Baraga County Michigan, USA as Fuzzy Sets. *Plant Ecology*, 134, 97-111.
- Bogárdi, I., Bárdossy, A., & Duckstein, L. (1990). Risk Management for Groundwater Contamination: Fuzzy Set Approach. In R. Khanpilvardi & T. Gooch (Eds.) *Optimizing the Resources for Water Management* (pp. 442-448). ASCE.
- Buckley, J. J., & Eslami E. (2002). *An Introduction to Fuzzy Logic and Fuzzy Sets*. Advances in Soft Computing. Physica-Verlag.
- Burrough, P. A. (1996). Natural Objects with Indeterminate Boundaries. In P. A. Burrough & A. U. Frank (Eds.), *Geographic Objects with Indeterminate Boundaries* (pp. 3-28). Taylor & Francis.
- Burrough, P.A., van Gaans, P.F.M., & Macmillan, R.A. (2000). High-Resolution Landform Classification Using Fuzzy k-Means. *Fuzzy Sets and Systems*, 113, 37-52.
- Cheng, T., Molenaar, M., & Lin, H. (2001). Formalizing Fuzzy Objects from Uncertain Classification Results. *International Journal of Geographical Information Science*, 15, 27-42.
- Clementini, E., & Di Felice, P. (1996a). A Model for Representing Topological Relationships between Complex Geometric Features in Spatial Databases. *Information Sciences*, 90(1-4):121-136.
- Clementini, E., & Di Felice, P. (1996b). An Algebraic Model for Spatial Objects with Indeterminate Boundaries. In P. A. Burrough & A. U. Frank (Eds.), *Geographic Objects with Indeterminate Boundaries* (pp. 153-169). Taylor & Francis.
- Clementini, E., & Di Felice, P. (2001). A Spatial Model for Complex Objects with a Broad Boundary Supporting Queries on Uncertain Data. *Data & Knowledge Engineering*, 37:285-305, 2001.
- Cohn, A. G., & Gotts, N. M. (1996). The 'Egg-Yolk' Representation of Regions with Indeterminate Boundaries. In P. A. Burrough & A. U. Frank (Eds.), *Geographic Objects with Indeterminate Boundaries* (pp. 171-187). Taylor & Francis.

- De Gruijter J., Walvoort, D., & Vangaans, P. (1997). Continuous Soil Maps—a Fuzzy Set Approach to Bridge the Gap between Aggregation Levels of Process and Distribution Models. *Geoderma*, 77, 169-195.
- Dilo, A., de By, R. A., & Stein, A. (2007). A System of Types and Operators for Handling Vague Spatial Objects. *International Journal of Geographical Information Science*, 21(4), 397-426.
- Dutta, S. (1989). Qualitative Spatial Reasoning: A Semi-Quantitative Approach Using Fuzzy Logic. In *1st International Symposium on the Design and Implementation of Large Spatial Databases* (pp. 345-364). LNCS 409, Springer Verlag.
- Dutta, S. (1991). Topological Constraints: A Representational Framework for Approximate Spatial and Temporal Reasoning. In *2nd International Symposium on the Design and Implementation of Large Spatial Databases* (pp. 161-180). LNCS 525, Springer Verlag.
- Edwards, G. (1994). Characterizing and Maintaining Polygons with Fuzzy Boundaries in GIS. In *6th International Symposium on Spatial Data Handling* (pp. 223-239).
- Egenhofer, M. J. (1994). Spatial SQL: A Query and Presentation Language. *IEEE Trans. on Knowledge and Data Engineering*, 6(1):86-94.
- Erwig, M., & Schneider, M. (1997). Vague Regions. In *5th International Symposium on Advances in Spatial Databases* (pp. 298-320). LNCS 1262, Springer Verlag.
- Finn, J. T. (1993). Use of the Average Mutual Information Index in Evaluating Classification Error and Consistency. *Int. Journal of Geographical Information Systems*, 7(4):349-366.
- Güting, R. H. (1988). Geo-Relational Algebra: A Model and Query Language for Geometric Database Systems. *International Conference on Extending Database Technology*, pp. 506-527.
- Hendricks Franssen, H., van Eijnsbergen, A. and Stein, A. (1997). Use of Spatial Prediction Techniques and Fuzzy Classification for Mapping Soil Pollutants. *Geoderma*, 77, 243-262.
- Kollias, V. J., & Voliotis, A. (1991). Fuzzy Reasoning in the Development of Geographical Information Systems. *International Journal of Geographical Information Systems*, 5(2), 209-223.
- Lagacherie, P., Andrieux, P., & Bouzigues, R. (1996). Fuzziness and Uncertainty of Soil Boundaries: From Reality to Coding in GIS. In P. A. Burrough & A. U. Frank (Eds.), *Geographic Objects with Indeterminate Boundaries* (pp. 275-286). Taylor & Francis.
- Liu, Y.-M., & Luo M.-K. (1997). *Fuzzy Topology. Advances in Fuzzy Systems — Applications and Theory*, vol. 9. World Scientific.
- Pauly, A., & Schneider, M. (2004). Vague Spatial Data Types, Set Operations, and Predicates. In *8th East-European Conference on Advances in Databases and Information Systems* (pp. 379-392).
- Pauly, A., & Schneider, M. (2005a). Identifying Topological Predicates for Vague Spatial Objects. *20th ACM Symposium on Applied Computing*, pp. 587-591.
- Pauly, A., & Schneider, M. (2005b). Topological Predicates between Vague Spatial Objects. In *9th International Symposium on Spatial and Temporal Databases* (pp. 418-432).
- Pauly, A., & Schneider, M. (2006). Topological Reasoning for Identifying a Complete Set of Topological Predicates between Vague Spatial Objects. In *19th International FLAIRS Conference* (pp. 731-736).
- Pawlak, Z. (1982). Rough Sets. Basic Notions. *International Journal of Computer and Information Science*, 11, 341-356.
- Petry, F. E., Cobb, M., Ali, D., Angryk, R., Paprzycki, M., Rahimi, S., Wen, L., & Yang, H. (2002). Fuzzy Spatial Relationships and Mobile Agent Technology in Geospatial Information Systems. In P.

- Matsakis & L.M. Sztandera (Eds.), *Soft Computing in Defining Spatial Relations* (pp. 123-155), volume in series: Soft Computing. Physica-Verlag.
- Schneider, M. (1996). Modelling Spatial Objects with Undetermined Boundaries Using the Realm/ROSE Approach. In P. A. Burrough & A. U. Frank (Eds.), *Geographic Objects with Indeterminate Boundaries* (pp. 141-152). Taylor & Francis.
- Schneider, M. (1997). *Spatial Data Types for Database Systems — Finite Resolution Geometry for Geographic Information Systems*. LNCS 1288, Springer-Verlag.
- Schneider, M. (1999). Uncertainty Management for Spatial Data in Databases: Fuzzy Spatial Data Types. In *6th International Symposium on Advances in Spatial Databases* (pp. 330-351). LNCS 1651, Springer Verlag.
- Schneider, M. (2000). Metric Operations on Fuzzy Spatial Objects in Databases. In *8th ACM Symposium on Geographic Information Systems* (pp. 21-26). ACM Press.
- Schneider, M. (2001a). A Design of Topological Predicates for Complex Crisp and Fuzzy Regions. In *20th International Conference on Conceptual Modeling* (pp. 103-116).
- Schneider, M. (2001b). Fuzzy Topological Predicates, Their Properties, and Their Integration into Query Languages. In *9th ACM Symposium on Geographic Information Systems* (pp. 9-14). ACM Press.
- Schneider, M. (2003). Design and Implementation of Finite Resolution Crisp and Fuzzy Spatial Objects. *Data & Knowledge Engineering*, 44(1), 81-108.
- Schneider, M., & Behr, T. (2006). Topological Relationships between Complex Spatial Objects. *ACM Transactions on Database Systems*, 31(1), 39-81.
- Shi W., & Guo W. (1999). Modeling Topological Relationships of Spatial Objects with Uncertainties. In: *International Symposium on Spatial Data Quality* (pp. 487-495).
- Shibasaki, R. (1993). A Framework for Handling Geometric Data with Positional Uncertainty in a GIS Environment. *GIS: Technology and Applications*, pp. 21-35, World Scientific.
- Tang X., & Kainz W. (2002). Analysis of Topological Relations between Fuzzy Regions in a General Fuzzy Topological Space. In: *Joint International Symposium on Geospatial Theory, Processing and Application*.
- Tilove, R. B. (1980). Set Membership Classification: A Unified Approach to Geometric Intersection Problems. *IEEE Transactions on Computers*, C-29, 874-883.
- Uery, E. L. (1996). A Conceptual Framework and Fuzzy Set Implementation for Geographic Features. In P. A. Burrough & A. U. Frank (Eds.), *Geographic Objects with Indeterminate Boundaries* (pp. 71-85). Taylor & Francis.
- Wang, F. (1994). Towards a Natural Language User Interface: An Approach of Fuzzy Query. *International Journal of Geographical Information Systems*, 8(2), 143-162.
- Wang, F., & Hall, G. B. (1996). Fuzzy Representation of Geographical Boundaries in GIS. *International Journal of Geographical Information Systems*, 10(5), 573-590.
- Wang, F., Hall, G. B., & Subaryono (1990). Fuzzy Information Representation and Processing in Conventional GIS Software: Database Design and Application. *International Journal of Geographical Information Systems*, 4(3), 261-283.
- Worboys, M. (1998). Computation with Imprecise Geospatial Data. *Computational, Environmental and Urban Systems*, 22(2):85-106.
- Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control*, 8, 338-353.

Zhan, B. F. (1998). Approximate Analysis of Topological Relations between Geographic Regions with Indeterminate Boundaries. *Soft Computing*, 2, 28-34.

Zinn, D., Bosch, J., & Gertz, M. (2007). Modeling and Querying Vague Spatial Objects Using Shapelets. *International Conference on Very Large Data Bases*, pp. 567-578.

Zhan, B.F. (1997). Topological Relations between Fuzzy Regions. In *ACM Symposium on Applied Computing* (pp. 192-196). ACM Press.

Zhan, B.F. (1998). Approximate Analysis of Topological Relations between Geographic Regions with Indeterminate Boundaries. *Soft Computing*, 2, 28-34.

KEY TERMS & DEFINITIONS

Abstract data type: A data type whose internals are hidden and whose components are only accessible through high-level methods.

Fuzzy spatial data type: A spatial data type either for fuzzy points, fuzzy lines, or fuzzy regions that are based on an infinite logic

Spatial data type: A data type like for representing geometries like points, lines, and regions in a database system.

Spatial vagueness: The feature of spatial objects that their boundary and/or interior is vague or indeterminate and cannot be precisely described.

Vague Spatial Algebra (VASA): A type system that comprises vague spatial data types, vague spatial operations, and vague spatial predicates.

Vague spatial data type: A spatial data type either for vague points, vague lines, or vague regions that are based on a three-valued logic with the truth values *true*, *false*, and *maybe*.

ACKNOWLEDGEMENTS

This work was partially supported by the National Science Foundation under grant number NSF-CAREER-IIS-0347574.